Exhibit 7

allowed us to type unknown individuals as *DRw52a* or *DRw52b* (ref. 22 and unpublished) in a considerable number of normal and diseased individuals. This form of analysis can be of considerable use in phylogenetic studies of human populations.

Taken together, these results can account for the evolution of the *DR* genes in the *DRw52* supertypic group (Fig. 4). The ancestral features of this family are the relatively recent duplication of the *βI* locus and the silencing of the *βII* locus by deletion of the first domain encoding exon[11]. The duplicated *DRβI* loci then diverged into *βI* and *βIII*. Further divergence resulted in a branching into two lineages (*DRw52a* and *DRw52b*) based on common alleles at the less polymorphic locus, *DRβIII*. In the *DRw52a* group, the *DRw6a* haplotype gave rise to the DR3 specificity by the gene conversion described here. The *DRw6b* haplotype was probably involved in an interchromosomal gene conversion with the *DR4 βIII* locus acting as donor.

This analysis provides a framework for assigning serological specificities to the products of the different loci of the *DRw52* haplotypes. Allelic differences in the product of the *DRβIII* locus split *DRw52* into *a* and *b*. It has already been shown that this locus encodes the DRw52 specificity for the case of the *DRw6b* haplotype[25]. We propose that the distinct epitopes DRw52a and DRw52b (Fig. 4) will correspond to serological and T-cell specificities. In addition to the product of locus *DRβIII*, each haplotype obviously also expresses the product of their *βI* locus, which determines the fine DR specificity.

The data described here represent an example of relatively rapid evolution of a multigene family in which the loci appear to diverge at different rates following a duplication event. The time of divergence may be estimated by analysing this group of haplotypes in other geographical (non-European) groups whose migratory patterns are known.

The divergence in this gene family is generated in part by gene conversion. Since this mechanism can involve the transfer of preselected epitopes, the resulting additional polymorphism is frequently maintained in the population. Therefore, even if gene conversion is a relatively rare event, it can play a major role in the generation of polymorphism by producing functionally effective variants. It is generally thought that this polymorphism confers a selective advantage to a population in terms of its ability to cope with various pathogens. A genetic system with multiple loci undergoing conversion events could regenerate polymorphism in populations which have undergone bottlenecks due to migration or adverse environmental factors.

We thank Chantal Mattmann and Madeleine Zufferey for technical assistance, Dr B. Schwartz for providing sequence data for *DR5* prior to publication and Dr J. J. van Rood for comments on the manuscript. This work was supported by a grant from the Swiss National Fund for Scientific Research.

Note added in proof: From a recent analysis of micropolymorphism of HLA-DR4 (ref. 26), we propose that one *DR4βI* allele, *Dw10*, has arisen by a gene conversion, with *DRw6aβI* acting as donor.

Received 17 January; accepted 4 April 1986.

1. Schwartz, R. *A. Rev. Immun.* 3, 237-261 (1985).
2. Nagy, Z. *et al. Immun. Rev.* 60, 61-83 (1981).
3. Kaufman, J. *et al. Cell* 36, 1-13 (1984).
4. Kaufman, J. F. & Strominger, J. L. *Nature* 297, 694-697 (1982).
5. Long, E. O. *et al. EMBO J.* 2, 389-394 (1983).
6. Gorski, J. *et al.* in *UCLA Symp.* 18, 47-56 (1984).
7. Speis, T. *et al. Proc. natn. Acad. Sci. U.S.A.* 82, 5165-5169 (1985).
8. Lee, J., Trowsdale, J. & Bodmer, W. F. *Proc. natn. Acad. Sci. U.S.A.* 79, 545-549 (1982).
9. Korman, A. *et al. Proc. natn. Acad. Sci. U.S.A.* 79, 1844-1848 (1982).
10. Wake, C. *et al. Proc. natn. Acad. Sci. U.S.A.* 79, 6979-6983 (1982).
11. Rollini, P., Mach, B. & Gorski, J. *Proc. natn. Acad. Sci. U.S.A.* 82, 7197-7201 (1985).
12. Tanigaki, N. & Tosi, R. *Immun. Rev.* 66, 5-37 (1982).
13. Angelini, G. *et al.* in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 582-584 (Springer, New York, 1984).
14. Hui, K. M. *et al.* in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 590-594 (Springer, New York, 1984).
15. Tieber, V. L. *et al. J. biol Chem.* 261, 2738-2742 (1986).
16. Kounlsky, P. *Biochimie* 65, 85-93 (1983).
17. Grosse-Wilde, H., Doxiadis, I. & Brandt, H. in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 249-264 (Springer, New York, 1984).
18. Denaro, M. *et al. EMBO J.* 3, 2029-2032 (1984).
19. Mengle-Gaw, L. *et al. J. exp. Med.* 160, 1184-1194 (1984).
20. Widera, G. & [illegible] R. A. *EMBO J.* 3, 1221-1225 (1984).
21. Cairns, J. S. *et al. Nature* 317, 166-168 (1985).
22. Angelini, G. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
23. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* 74, 560-564 (1977).
24. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* 74, 5463-5467 (1977).
25. Gorski, J. *et al. J. exp. Med.* 162, 105-116 (1985).
26. Gregersen, P. K. *et al. Proc. natn. Acad. Sci. U.S.A.* 83, 2642-2646 (1986).

Hepatitis B virus DNA integration in a sequence homologous to v-*erb*-*A* and steroid receptor genes in a hepatocellular carcinoma

Anne Dejean*, Lydie Bougueleret†, Karl-Heinz Grzeschik‡ & Pierre Tiollais*

* Unité de Recombinaison et Expression Génétique (INSERM U.163, CNRS UA.271) and † Unité d'Informatique Scientifique, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France
‡ Institut für Humangenetik der Westfälischen Wilhems-Universität, 4400 Munster, FRG

Hepatitis B virus (HBV) is clearly involved in the aetiology of human hepatocellular carcinoma (HCC)[1] and the finding of HBV DNA integration into human liver DNA in almost all HCC studied[2-7] suggested that these integrated viral sequences may be involved in liver oncogenesis. Several HBV integrations in different HCCs[8,9] and HCC-derived cell lines[10-14] have been analysed after molecular cloning without revealing any obvious role for HBV. From a comparison of a HBV integration site present in a particular HCC[8] with the corresponding unoccupied site in the non-tumorous tissue of the same liver, we now report that HBV integration places the viral sequence next to a liver cell sequence which bears a striking resemblance to both an oncogene (v-*erb*-*A*) and the supposed DNA-binding domain of the human glucocorticoid receptor and human oestrogen receptor genes. We suggest that this gene, usually silent or transcribed at a very low level in normal hepatocytes, becomes inappropriately expressed as a consequence of HBV integration, thus contributing to the cell transformation.

We have previously reported the molecular cloning of the single integrated viral sequence present in the liver tumorous nodule of patient D and we have determined the sequences of the cellular-viral junctions[8]. The viral insertion was a continuous subgenomic fragment 1.4 kilobases (kb) long (Fig. 1*a*) containing the cohesive-end region, gene *C* and the beginning of gene *pre-S1*. We therefore used the 1.1-kb and 5.8-kb *Hin*dIII cellular fragments and the 1.8-kb *Eco*RI host-viral fragment (respectively referred to as LT (left tumour), RT (right tumour) and MT (medium tumour); Fig. 1*a*) to isolate the unoccupied site from a λ phage library of DNA extracted from the non-tumorous liver of patient D. This part of the liver did not seem to contain any integrated HBV sequences. Seven overlapping clones, hybridizing to one and/or the other of the three probes, were isolated and represented 32 kb of cellular DNA at the unoccupied site (Fig. 1*b*). Southern blots of restriction digests of the seven clones using total human DNA as a probe showed that the host sequence at the viral insertion site corresponds mainly to unique sequence DNA (Fig. 1*a*, *b*, solid bars). Comparison between the restriction maps of the unoccupied site (Fig. 1*b*) and the integrated site (Fig. 1*a*) did not reveal any major genomic rearrangements in the cellular DNA. Integration took place within a small *Eco*RI fragment of 400 base pairs (bp) which we subcloned and refer to as MNT (medium non-tumour) (Fig. 1*b*).

To investigate whether HBV became integrated in the vicinity of a cellular gene in the human genome, we determined the nucleotide sequence[15] of the normal allele. This sequence

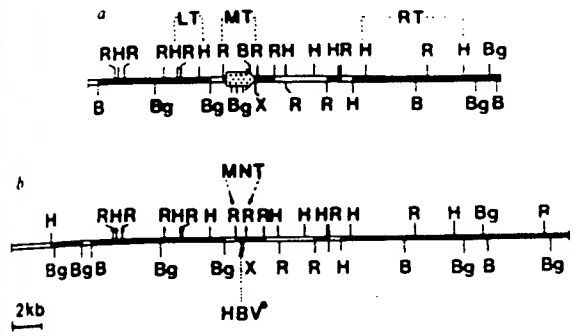

Fig. 1 *a*, Restriction map of the occupied site. This has previously been isolated from a library of cellular DNA extracted from the tumorous part of the liver of patient D[8]. The stippled region represents the 1.4-kb integrated HBV sequence. The arrowhead denotes the same orientation as the viral (+)strand[30]. Solid bars denote unique sequences and open bars regions containing repetitive cellular sequences. Restriction sites are: R, *Eco*RI; B, *Bam*HI; Bg, *Bgl*I; H, *Hin*dIII; X, *Xho*I. The cloned 1.1-kb *Hin*dIII, 1.8-kb *Eco*RI and 5.8-kb *Hin*dIII cellular fragments are referred to as LT, MT and RT. *b*, Restriction map of the unoccupied site. The cloning experiment in which partial *Mbo*I digests of cellular DNA were cloned into λ L47.1 was as described previously[11]. A library, made with cellular DNA from the non-tumorous part of the liver of patient D, was screened with LT, MT and RT as probes. Seven positive overlapping clones were analysed, representing 32 kb of cellular DNA. The cloned 400-bp *Eco*RI cellular fragment is referred to as MNT. The site of HBV integration within MNT is indicated by an arrow.

revealed the presence of an open reading frame (ORF) of 519 nucleotides which was interrupted in the middle by the viral insertion (Fig. 2). Since methionine codons were present only at the 3' end of the ORF, this region can only correspond to a single exon from a split gene. Two acceptor-like (A-L1, A-L2) and two donor-like (D-L1, D-L2) consensus sequences of splice junctions[16] could be localized, respectively, at the 5' end and the 3' end of the ORF (Fig. 2). A computer-assisted analysis[17] showed that the region between A-L2 and D-L1 had a strong probability of being an exon. A search for related sequences in the NBRF protein data bank[18] identified a remarkable homology (22 identities out of 49) between the translation product of the exon-like region of the ORF (amino-acid residues 69–117) and the amino-terminal region of the v-*erb-A* oncogene product (residues 8–58)[19] (Fig. 3). Moreover, a significant homology has recently been reported between the v-*erb-A* protein and either the human glucocorticoid receptor (hGR)[20] or the human oestrogen receptor (ER)[21]. The alignment of the amino-acid sequence of hGR and ER with the exon-like protein product showed, in both cases, 19 identities out of 49 (Fig. 3). As already observed for v-*erb-A*, hGR[20] and ER[21], the identity became greater beyond amino acid 96 of the ORF, which corresponds to a cysteine-rich region. In particular, the four cysteines conserved between v-*erb-A*, hGR and ER are present at the same position in the ORF. Although the homology continues for v-*erb-A*, hGR and ER—all derived from cDNA clones—no significant homology was found beyond residue 117 for the ORF. However, since the D-L1 sequence (next to residue 117) is strictly identical to the consensus sequence of a donor site, it could thus correspond to an exon/intron boundary.

The integration of HBV sequences interrupted the cellular open reading frame and generated a microdeletion of 7–12 bp (boxed in Fig. 4). This minor rearrangement provides evidence that the situation we are studying in patient D is probably very near the initial integration event. In addition to the microdeletion, the viral integration—interrupting the cellular ORF—generated a new viral-host hybrid sequence such that the first 39 codons of the viral *pre-S1* gene became fused and in phase with the last 28 codons of the cellular exon (Fig. 4). Remarkably,

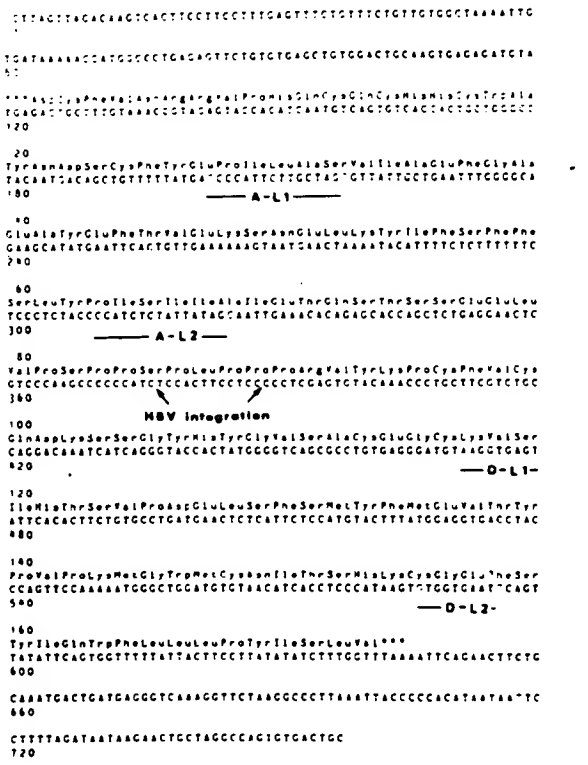

Fig. 2 Nucleotide sequence of the unoccupied site. Nucleotides are numbered at the left side. The deduced amino-acid sequence of the 519-bp open reading frame is shown above the nucleotide sequence. The amino-acid sequence is numbered from the first codon of the ORF. A large number of splice junction sequences have been reported[16]. The compilation of the data supports the consensus $^{(T)}_{(C)}{}_{11}N^{C}_{T}AG/G$ for acceptors and the consensus $^{C}_{A}AG/GT^{A}_{G}AGT$ for donors. The two acceptor-like (A-L1 and A-L2) and donor-like (D-L1 and D-L2) sequences are underlined. The site of HBV integration in the middle of ORF is indicated by an arrow. The cloned 3.4-kb *Hin*dIII fragment, encompassing the unintegrated site in the normal allele, was sonicated, treated with the Klenow fragment of DNA polymerase plus deoxyribonucleotides (2 h, 15 °C) and fractionated by agarose gel electrophoresis. Fragments of 400–700 bp were excised and electroeluted. DNA was ethanol-precipitated, ligated to dephosphorylated *Sma*I-cleaved M13 mp8 replication form DNA and transfected into *Escherichia coli* strain TG-1 by the high-efficiency technique of Hanahan[31]. Recombinant clones were detected by plaque hybridization using the MNT (Fig. 1*b*) subclone DNA as a probe. Single-stranded templates were prepared from plaques exhibiting positive hybridization signals and were sequenced by the dideoxy chain termination procedure[15] using buffer gradient gels[32].

the viral genome became integrated a few nucleotides upstream from the most conserved Cys-rich portion of the ORF (Fig. 3), maintaining the integrity of this region.

Using a panel of 17 mouse-human and Chinese hamster-human somatic cell hybrid DNAs[22,23], we localized the ORF to chromosome 3 (data not shown), while the c-*erb-A* oncogene[24], hGR[25] and ER[26] have been mapped, respectively, to human chromosomes 17, 5 and 6. In preliminary experiments, we hybridized the MNT probe, encompassing the exon-like region of ORF, to a Northern blot of polyadenylated RNAs extracted from five human livers, but found no detectable transcripts. A large number of human fetal and adult tissues will have to be tested similarly to reveal any active transcription of this region.

The conserved Cys-rich region which extends over 60 amino-acid residues in v-*erb-A* protein, hGR and ER is thought to include the DNA-binding domain of the molecule[20,21]. We can thus speculate that the corresponding homologous region of ORF, truncated by the exon–intron boundary, is part of a cellular gene that shares a common functional domain with hGR, ER

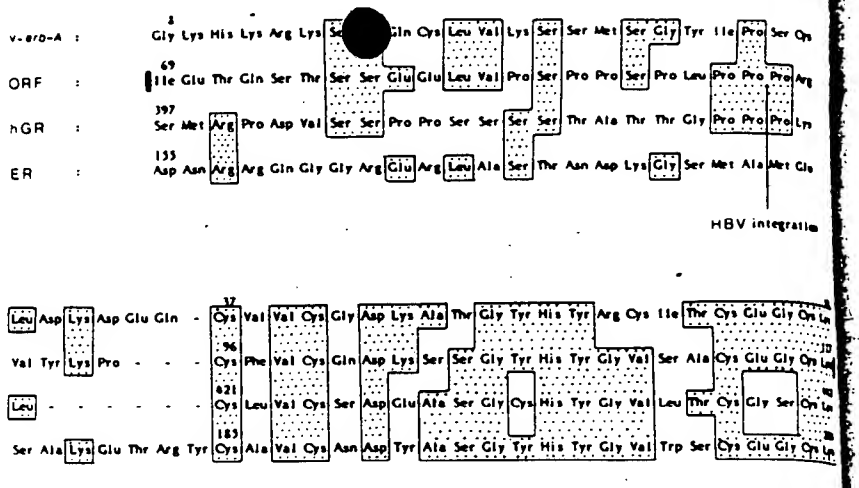

Fig. 3 Amino-acid sequence alignment of the v-*erb-A* oncogene protein[19], translation product of the exon-like region of ORF, the human glucocorticoid receptor (hGR)[20] and the oestrogen receptor (ER)[21]. The limits of the exon-like region of ORF, defined by A-L2 and D-L1 boundaries, are indicated by rectangles. To predict the location of exon-like regions, we used the discriminating program PREDICTOR[17]. Two subsets of the GenBank data library, containing either only exon or only intron sequences, were taken as reference pool. The program PROBE3-EXPLOR3 (ref. 18), allowing the search for ambiguous nucleic or peptidic patterns, was used to screen both the NBRF (proteins) and the GenBank (DNA) data banks. These programs were run on a MV8000 32-bit minicomputer. Amino-acid residues 69–117 from the ORF were aligned with amino-acid residues 8–58 from $p75^{gag\text{-}erb\text{-}A}$, residues 397–442 from hGR and residues 115–206 from ER. Identical residues are boxed and gaps are indicated by dashes. The HBV integration site, upstream from the cysteine-rich region, is indicated.

and v-*erb-A* gene products and which could exert a transcriptional regulatory function on specific genes.

Although the way in which HBV participates in the formation of a liver cancer is unknown, the experiments reported here could promote our understanding of one possible mechanism of HBV carcinogenesis. In patient D the viral integration, interrupting the exon-like region (Fig. 2), created a chimaeric viral-host open reading frame (Fig. 4). The HBV insertion took place a few nucleotides upstream from the beginning of the putative DNA-binding domain. Since a viral promoter has been defined by *in vitro* transcription approximately 30 nucleotides upstream from the initiator codon of the *pre-S1* gene[27], we suggest that, in the tumorous part of the liver, a readthrough transcription occurred from the viral promoter. Although protein or RNA from the tumour is no longer available to test this hypothesis, it is most probable that inappropriate activation of the putative gene as a consequence of HBV integration resulted in expression of a truncated protein at greater levels than that of the native protein. This protein could participate directly in the subsequent cell transformation.

Several arguments suggest that hormonal factors are involved in human hepatocarcinogenesis. The incidence of HCC is three- to sixfold greater in males[1], and the use of oral contraceptives in females is associated with the development of hepatic adenoma[28]. Moreover, the ability of oestrogenic hormones to function as promoters of neoplastic development in rat liver has been clearly demonstrated[29]. The finding that HBV sequences have become integrated into a putative cellular gene sharing homology with the steroid receptor genes is therefore intriguing and suggests that, in some cases, hormonal and HBV carcinogenesis may be directly related.

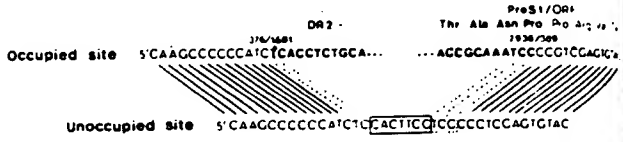

Fig. 4 DNA sequences at the HBV integration site. The sequences of the left and right host-viral DNA junctions at the occupied site (upper sequence) are compared with the human DNA sequence at the unoccupied site (lower sequence). The bold-face letters indicate the viral sequence. Nucleotides of the ORF and the HBV genome[33] are numbered. Homologous nucleotides between the two sequences are indicated by sloping lines. The 7-bp CACTTCC present in the normal allele and deleted after the viral integration in the occupied site is boxed. Because HBV DNA and cellular DNA shared a 2-bp and 3-bp sequence homology at a point coincident, respectively, with the left and right host-viral junctions (dashed lines), the deleted fragment could be up to 12 bp long. The DR2 copy of the 11-bp viral direct repeat specifically involved in HBV integration[8] is indicated. The putative chimaeric protein, generated by the HBV inversion, between the first 29 amino acids of the viral *pre-S1* gene product and the last 28 amino acids of the cellular exon protein product is partially represented at the fusion site.

We thank Drs Christian Bréchot, Pierre Sonigo, Simon Wain-Hobson and Jean Weissenbach for advice and discussion and Louise-Marie Da for typing the manuscript. This work was supported by NIH grant CA97300-02, the Economic European Community and the Faculté de Médecine Lariboisière Saint-Louis (University of Paris VII).

Received 5 February; accepted 16 April 1986.

1. Beasley, R. P. & Hwang, L. Y. in *Viral Hepatitis and Liver Disease* (eds Vyas, G. N., Dienstag, J. L. & Hoofnagle, J. H.) 209–214 (Grune & Stratton, New York, 1984).
2. Charkraborty, P. R., Ruiz-Opazo, N., Shouval, D. & Shafritz, D. A. *Nature* **286**, 531–533 (1980).
3. Bréchot, C., Pourcel, C., Louise, A., Rain, B. & Tiollais, P. *Nature* **286**, 533–535 (1980).
4. Edman, J. C., Gray, P., Valenzuela, P., Rall, L. B. & Rutter, W. J. *Nature* **286**, 535–538 (1980).
5. Bréchot, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3906–3910 (1981).
6. Shafritz, D. A., Shouval, D., Sherman, H. I., Hadziyannis, S. J. & Kew, M. C. *New Engl. J. Med.* **305**, 1067–1073 (1981).
7. Chen, D. S., Hoyer, B. H., Nelson, J., Purcell, R. H. & Genn, J. L. *Hepatology* **2**, 42S–45S (1982).
8. Dejean, A., Sonigo, P., Wain-Hobson, S. & Tiollais, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5350–5354 (1984).
9. Rogler, C. E. *et al. Science* **230**, 319–322 (1985).
10. Koshy, R. *et al. Cell* **24**, 215–223 (1983).
11. Dejean, A., Bréchot, C., Tiollais, P. & Wain-Hobson, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2505–2509 (1983).
12. Shaul, Y. *et al. J. Virol.* **51**, 776–787 (1984).
13. Mizusawa, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 208–212 (1985).
14. Yaginuma, K., Kobayashi, M., Yoshida, E. & Koike, K. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4458–4462 (1985).
15. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
16. Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
17. Claverie, J. M. & Bougueleret, L. *Nucleic Acids Res.* **14** (in the press).
18. Claverie, J. M. & Sauvaget, I. *Cabios* **1**, 95–104 (1985).
19. Debuire, B. *et al. Science* **224**, 1456–1459 (1984).
20. Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Nature* **318**, 670–677 (1985).
21. Green, S. *et al. Nature* **320**, 134–139 (1986).
22. Johannsmann, R., Heilkuhl, B. & Grzeschik, K. H. *Hum. Genet.* **56**, 361–363 (1981).
23. Huerre, C. *et al. Ann. Genet.* **27**, 5–10 (1984).
24. Spurr, N. K. *et al. EMBO J.* **3**, 159–163 (1984).
25. Hollenberg, S. M. *et al. Nature* **318**, 635–641 (1985).
26. Walter, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7889–7893 (1985).
27. Rall, L. B., Standring, D. N., Laub, O. & Rutter, W. J. *Molec. cell. Biol.* **3**, 1766–1773 (1983).
28. Edmonson, M. A., Henderson, B. & Benton, B. *New Engl. J. Med.* **294**, 470 (1976).
29. Wanless, I. R. & Medline, A. *Lab. Invest.* **46**, 313–320 (1982).
30. Tiollais, P., Pourcel, C. & Dejean, A. *Nature* **317**, 489–495 (1985).
31. Hanahan, D. *J. molec. Biol.* **166**, 557–580 (1983).
32. Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963–3965 (1983).
33. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646–650 (1979).